Extending the Farkas Lemma Approach
to Necessity Conditions
to Infinite Programming

R.A. Tapia

M.W. Trosset

			Form Approved OMB No. 0704-0188
	Report Documentation Page		

1. REPORT DATE **05 MAY 1993**	2. REPORT TYPE	3. DATES COVERED **00-00-1993 to 00-00-1993**

4. TITLE AND SUBTITLE **Extending the Farkas Lemma Approach to Necessity Conditions to Infinite Programming**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**31**	

Extending the Farkas Lemma Approach to Necessity Conditions to Infinite Programming

R.A. Tapia [*]   M.W. Trosset [†]

May 5, 1993

1

Abstract

Under mild assumptions, the classical Farkas lemma approach to Lagrange multiplier theory is extended to an infinite programming formulation. The main result generalizes the usual first-order necessity conditions to address problems in which the domain of the objective function is Hilbert space and the number of constraints is arbitrary. The result is used to obtain necessity conditions for a well-known problem from the statistical literature on probability density estimation.

Key words: Lagrange multiplier theory, Farkas lemma, infinite programming, mathematical programming.

1 Introduction

In 1951, Kuhn and Tucker [14] developed a Lagrange multiplier theory for mathematical programming problems that contain inequality constraints. In this theory, the domain of the objective function is Euclidean space and the constraint functionals are indexed by a finite set. Their development, explicated and popularized by Fiacco and McCormick ([6], Chapter 2), invokes the classical Farkas lemma (Farkas [5]) to generate a vector (the Lagrange multipliers) that can be viewed as a "weighting" of the finite set of constraints. For years we have taught this material, each time pondering the extent to which this development of Lagrange multiplier theory depends on the finite dimensionality of Euclidean space and the finiteness of the constraint set. Somewhat recently, our interest was enhanced when we learned of an interesting infinite programming problem in the statistics literature for which we could state a formal generalization of the usual first-order necessity conditions, with no known theoretical justification for doing so. In the present study of Lagrange multiplier theory, we have not only succeeded in generalizing the Farkas lemma approach, but have also acquired new insight into the essential features of that approach.

Of late, it has become fashionable to refer to the first-order necessity conditions as the "Karush-Kuhn-Tucker conditions," rather than the "Kuhn-Tucker conditions." This observation motivated us to carefully study the origins of these conditions, resulting in a fascinating excursion into the history of nonlinear programming and the classical calculus of variations. Because of the pedagogical nature of the present work, we believe that it is appropriate to share what we have learned. In passing, we note that Prékopa [17] effectively argues that fundamental ideas concerning the optimality conditions for nonlinear programming can be found in some early papers in mechanics by Fourier, Cournot, and Farkas and also by Gauss, Ostrogradsky and Hamel.

In the 1930s, there flourished at the University of Chicago a school of thought in the calculus of variations that was founded by G. A. Bliss. Researchers associated with this school included L. M. Graves, H. H. Goldstine, M. R. Hestenes, A. S. Householder, W. Karush, E. J. McShane, W. T. Reid, F. A. Valentine, and many others. Lagrange multiplier theory for the equality constrained mathematical programming problem was known, in one form or another, to most researchers in the classical calculus of variations, who were fully aware that it could be derived from various theories for more general

problems. Bliss [3] presented an elegant exposition of this theory in the first section of a 1938 survey of normality and abnormality in the calculus of variations. The first section of Bliss's article is entitled, "Abnormality for minima of functions of a finite number of variables," and it begins, "The significance of the notion of abnormality in the calculus of variations can be indicated by a study of the theory of the simpler problem of finding ...." (p. 367). What is particularly enlightening about this section is that it reveals how the Chicago school regarded finite-dimensional problems: in a pre-computational era, the theoretically less challenging case of finite dimensions was primarily valued as a training ground for developing intuition about more difficult problems.

Another interest of the Chicago school was the incorporation of inequality constraints. They often used squared slack variables to extend known theory and insight from the more standard equality constrained problem to the less standard inequality constrained problem. Hestenes recalled that this device, as well as the techniques of successive linear and quadratic programming (techniques usually attributed to the post-war mathematical programming community), was a standard tool of the Chicago school.

In his 1937 Ph.D. thesis, Valentine [23] studied the problem of Lagrange with differential inequality constraints of the form

$$g(x, y, dy/dx) \geq 0.$$

He replaced this constraint with the equality constraint

$$g(x, y, dy/dx) = (dz/dx)^2,$$

where $z(x)$ is an auxiliary function satisfying a particular initial condition, and applied the standard theory. Today, many authors refer to the use of squared slack variables as the "Method of Valentine." The first author of the present paper, undoubtedly influenced by the instruction of Hestenes and Valentine in his graduate education at UCLA, often employs this method to develop insight in elementary courses. For example, a quick way to derive Lagrange multiplier theory for general nonlinear programming is to add squared slack variables to the inequality constraints, then apply the standard theory for equality constrained nonlinear programming. This approach does not establish the nonnegativity of the multipliers of the inequality constraints; however, their nonnegativity follows directly from the second-order

4

necessity conditions. Collectively, then, one obtains the same first-order and second-order necessity conditions from this elementary approach. The price that one pays is that regularity (linear independence of the gradients of the binding constraints) must be assumed, so that it is not possible to reap the benefits of more subtle constraint qualifications.

Given these interests of the Chicago school, it made perfect sense for Graves to assign to Karush, as a topic suitable for his master's thesis, the simple problem of extending Bliss's *finite-dimensional* treatment of equality constraints to the case of inequality constraints. Karush [12] handled his assignment beautifully, deriving first-order necessity conditions using the Farkas lemma and formally stating the Kuhn-Tucker constraint qualification as "Property Q." Thus, Karush's 1939 thesis contains the Kuhn-Tucker theory *in all of its particulars*. The constraint qualification Q was implicit in Bliss [3], who showed that it was implied by regularity, which he then assumed. Karush, however, assumed only the constraint qualification, according it a privileged status that Bliss had not.

One characteristic of the usual (Kuhn-Tucker) approach to Lagrange multiplier theory is that it requires the multiplier of the gradient of the objective function (say λ_0) to be unity. This requirement is what necessitates a constraint qualification: it is generally known that this hypothesis is unnecessary if λ_0 is allowed to vary freely. For nonlinear programming with inequality constraints, this fact is usually attributed to John [11] and called the Fritz John condition. (The device itself was first studied by Mayer [15]; in fact, this is the standard formulation of Lagrange multiplier theory in the calculus of variations.) It is of value, here, to point out that Karush also presented a form of the John theory. A straightforward application of Bliss's treatment of the equality constrained problem to the squared slack variable formulation of the inequality constrained problem led him to a result that is identical to John's, but without the sign restrictions on the multipliers. Karush then observed that the sign restrictions followed from Bliss's second-order necessity theory. As we have already observed, however, the price of this derivation is the assumption of regularity and two continuous derivatives. Accordingly, Karush employed this approach in a manner that we have attributed to the Chicago school, viz. to obtain insights that led him to the Karush-Kuhn-Tucker theory.

Karush was never encouraged to publish his work, presumably because the finite-dimensional case was then deemed too elementary to be of inde-

5

pendent interest. It remained virtually unknown until Kuhn discovered a reference to it in a 1974 textbook by Takayama [19] and obtained a copy of the thesis. Kuhn attempted to "set the record straight" at a 1975 AMS symposium (Kuhn [13]), and even went so far as to refer to the "Karush conditions." The same year, Hestenes [10], who had directed Karush's Ph.D. thesis, noted Karush's work in his book on optimization theory. Even so, it is only very recently that Karush's work has been generally acknowledged.

In summary, our historical investigations have led us to strongly support the "Karush-Kuhn-Tucker" terminology. There is no question of the importance of the Kuhn-Tucker paper in the history of mathematical programming, but there is also no question that Karush obtained the identical result twelve years earlier. Furthermore, not only is Karush himself deserving of recognition, but we believe that the use of his name is a fitting tribute to the members of the great Chicago school of Bliss, whose deep understanding of mathematical programming has not been properly recognized and appreciated. Because it was not until the 1950s that there was a demand for the finite-dimensional theory, the Chicago researchers were simply two decades ahead of their time. The Karush-Kuhn-Tucker conditions represent a rare instance in which it is possible to document just how much of mathematical programming these researchers understood and anticipated.

Returning to the present paper, our objective is to extend the classical Farkas lemma approach to mathematical programming problems in which the domain of the objective function is Hilbert space and the constraint functionals are indexed by an arbitrary set. Our approach carefully mimics the finite programming development. It is based on a generalized Farkas lemma, and replaces the Lagrange multiplier vector with a measure on the (possibly infinite) index set. If this measure is absolutely continuous, then it can be represented as a (density) function on the index set. Because our point of view may seem unnatural to some readers otherwise familiar with Lagrange multiplier theory, we briefly digress to motivate it.

Consider vectors $x_1, \ldots, x_k \in R^n$, scalar weights $u_1, \ldots, u_k \in R^n$, and the weighted sum

$$\sum_{i \in I} u_i x_i ,$$

where the index set $I = \{1, \ldots, k\}$. By defining a measure μ on the Borel sets of R^n that concentrates on $\{x_1, \ldots, x_k\}$ and satisfies $\mu(\{x_i\}) = u_i$, we

can write

$$\sum_{i \in I} u_i x_i = \int_{\mathrm{R}^n} x \mu(dx) \ .$$

Thus, a set of weights can be viewed as a measure and a weighted sum can be viewed as a (Lebesgue) integral with respect to that measure. When the weights are nonnegative and sum to unity, μ is a probability measure and probabilists call the integral an expectation.

Now consider the index map $i \mapsto x_i$, which embeds I in R^n. The measure μ induces a measure u on the subsets of I by $u(\{i\}) = \mu(\{x_i\})$. This allows us to further write

$$\sum_{i \in I} u_i x_i = \int_{\mathrm{R}^n} x \mu(dx) = \int_I x_i u(di) \ ;$$

hence, our point of view that a set of weights is a measure on an index set. It is this perspective that will lead to a manageable statement of generalized first-order conditions.

The flavor of our generalization of Lagrange multiplier theory is not entirely new. Semi-infinite programming is also concerned with problems in which the constraint functionals are indexed by an infinite set, although the domain of the objective function is still assumed to be Euclidean space. The famous paper by John [11] posed a semi-infinite programming problem; however, John exploited the finite dimensionality of R^n to reduce the number of constraints to $n + 1$. More recently, a multiplier theorem of precisely the sort that we seek was obtained by Goberna, López, and Pastor [8]. The authors use a generalized Farkas lemma and retain the full set of constraints; however, their result also depends critically on the finite dimensionality of Euclidean space.

It should be noted that a number of authors have published multiplier theorems in very abstract settings. The standard formulation is that of Guignard [9], who derived both necessity and sufficiency conditions for the problem

$$
\begin{array}{ll}
\text{maximize} & \psi(x) \\
\text{subject to} & x \in C \subset X \\
& a(x) \in B \subset Y \ ,
\end{array}
$$

where X and Y are real Banach spaces and $\psi : X \to (-\infty, +\infty)$ and $a : X \to Y$ are Fréchet differentiable. Guignard's multiplier is an element of the

7

topological dual space of Y, and her entire approach is markedly different from ours.

The primary purpose of the present paper is pedagogical. That is, we wish to demonstrate that by (i) interpreting the vector of Lagrange multipliers as a measure on the index set of constraints and by (ii) utilizing tools from functional analysis and probability theory, the standard finite-dimensional approach to multiplier theory (Karush-Kuhn-Tucker) can be successfully generalized to infinite programming in Hilbert space. This exercise, however, is not entirely pedagogical, for we also believe that there are important infinite programming problems to which our theory can be profitably applied. Therefore, after in Section 2 deriving first-order necessity conditions for general infinite programming problems, in Section 3 we will consider results that facilitate the use of these conditions. In Section 4, by way of an example, we will also apply this theory to obtain necessity conditions for a constrained optimization problem from the statistical literature on probability density estimation. However, we have deferred to another paper (Trosset [21]) an investigation of the statistical consequences of these conditions.

2 Main Theorem

We begin with a real Hilbert space X with inner product $\langle \cdot , \cdot \rangle$. By the general nonlinear programming problem — problem (NLP) for short — we mean the constrained optimization problem

$$
\begin{aligned}
\text{maximize} \quad & f(x) \\
\text{subject to} \quad & g_\alpha(x) \geq 0 \quad && \forall \alpha \in I \\
& h_\beta(x) = 0 \quad && \forall \beta \in J ,
\end{aligned}
$$

where $f, g_\alpha, h_\beta : X \to (-\infty, +\infty)$. We assume that the index sets I and J have corresponding sigma fields $\mathbf{I}$ and $\mathbf{J}$ such that the pairs $(I, \mathbf{I})$ and $(J, \mathbf{J})$ are measure spaces. Measures on these spaces will be denoted by t, u, λ, etc. At times, we will also endow I and J with topologies. Typically, I and J will be subsets of Euclidean space. For each $x \in X$, we define the index subset $I_0(x) := \{\alpha \in I : g_\alpha(x) = 0\}$.

We assume that $f, g_\alpha, h_\beta \in C^1(X)$. For each $x \in X$, the sets $\nabla A_0(x) := \{\nabla g_\alpha(x) : \alpha \in I_0(x)\}$ and $\nabla B(x) := \{\nabla h_\beta(x) : \beta \in J\}$ are assumed to be

8

Borel measurable. We also assume that the index maps $\alpha \mapsto \nabla g_\alpha(x)$ and $\beta \mapsto \nabla h_\beta(x)$ are Borel bimeasurable functions. This will enable measures on I and J to induce measures on X, and also conversely. Measures on X will be denoted ∇F, ∇G, etc.

For technical reasons, we will sometimes further assume that the functions g_α and h_β are elements of a real Hilbert space Γ. In that event, we will assume that the sets $A_0(x) := \{g_\alpha : \alpha \in I_0(x)\}$ and $B := \{h_\beta : \beta \in J\}$ are Borel measurable. We will also assume that the index maps $\alpha \mapsto g_\alpha$ and $\beta \mapsto h_\beta$ are Borel measurable functions. This will enable measures on I and J to induce measures on Γ. Such measures will be denoted by F, G, etc.

In this section we will derive necessary conditions for a point $x^* \in X$ to be a local solution of problem (NLP). To do so, we modify and generalize Fiacco's and McCormick's [6] presentation of the first-order theory for the finite-dimensional case. The key to this generalization is the concept of the expectation of a measure on a Hilbert space. Toward this end, in what follows H will denote a real Hilbert space with inner product $(\cdot, \cdot)$. Following Parthasarathy ([16], Definition 3.2, p. 168) we make the following definition.

Definition 2.1 *Let μ be a measure on H. If the linear functional $L(y) := \int (y, x)\mu(dx)$ is continuous, then the expectation of μ, which we denote by $\int x\mu(dx)$, is defined to be the Riesz representer of L.*

At this point it will be of value to introduce some basic notation. Let $\mathbf{M}(K)$ denote the family of totally finite positive measures that concentrate on the set $K \subset H$, and let $\mathbf{M}_1(K)$ denote the family of probability measures that concentrate on the set $K \subset H$. We are interested in the sets of expectations

$$C(K) = \left\{ \int x\mu(dx) : \mu \in \mathbf{M}(K) \right\}$$

and

$$C_1(K) = \left\{ \int x\mu(dx) : \mu \in \mathbf{M}_1(K) \right\} .$$

The set $C_1(K)$ is essentially the convex hull of K, and the set $C(K)$ is essentially the cone generated by $C_1(K)$. It should be clear that $C(K)$ and $C_1(K)$ are convex. In the next section we will demonstrate that $C_1(K)$ is also compact. The closedness of $C(K)$ will be of fundamental importance in our theory. In the next section we will construct a condition which guarantees

that $C(K)$ is closed. However, for the moment we will assume that it is closed.

We now generalize a famous result.

Lemma 2.1 *(Generalized Farkas Lemma): Let H denote a real Hilbert space with inner product $(\cdot\,,\cdot)$. Let $x_0 \in H$ and $K \subset H$. Assume that $C(K)$ is closed. Then the following are equivalent:*

(i) $\quad \forall\, y \in H$, $\quad (y,x) \geq 0 \quad \forall\, x \in K \quad$ *entails* $\quad (y,x_0) \geq 0$;

(ii) $\quad \exists\, \mu \in \mathbf{M}(K) \quad$ *such that* $\quad x_0 = \displaystyle\int x\mu(dx)$.

Proof: We utilize the notion of a dual cone, introduced by Dieudonné [4] in his proof of the Hahn-Banach theorem. The dual cone C^* of a cone C is the set of all continuous linear functionals nonnegative on C.

Consider (i). If $(y,x) \geq 0 \;\forall x \in K$, then $(y,\int x\mu(dx)) = \int (y,x)\mu(dx) \geq 0 \;\forall \mu \in \mathbf{M}(K)$, i.e. $y \in C(K)^*$. Hence, (i) is equivalent to the assertion that $y \in C(K)^*$ entails $(y,x_0) \geq 0 \;\forall y \in H$, or simply that $x_0 \in C(K)^{**}$. By Lemma 5.6 in Girsanov [7], $C(K)^{**}$ is the weak closure of the convex hull of $C(K)$. Since $C(K)$ is closed and convex, it follows that (i) is equivalent to $x_0 \in C(K)^{**} = C(K)$. But (ii) is a direct statement that $x_0 \in C(K)$; hence, (i) and (ii) are equivalent. $\qquad\square$

Remark: It is possible to give an elementary, but more complicated proof of this result. The very elegant proof that we have presented was suggested to us anonymously by the referee. This is an amusing realization of Valentine's [22] admonition to "always look at the dual situation when working with convex sets for it may save you some embarrassment."

Associated with problem (NLP) is the generalized Lagrangian gradient

$$\ell'(x,u,\lambda) := \nabla_x f(x) - \int_I \nabla_x g_\alpha(x)u(d\alpha) + \int_J \nabla_x h_\beta(x)\lambda(d\beta) \,,$$

which is guaranteed to exist if the sets $\nabla A(x)$ and $\nabla B(x)$ are compact and the measures u and λ are totally finite. Our goal is to derive necessary conditions for solving problem (NLP) that involve this expression. We are now in a position to characterize some of these conditions.

Suppose that x is a feasible point of problem (NLP). Let

$$Z_1(x) := \{z \in X : \langle z, \nabla g_\alpha(x)\rangle \geq 0 \quad \forall\, \alpha \in I_0(x),$$
$$\langle z, \nabla h_\beta(x)\rangle = 0 \;\; \forall\, \beta \in J, \quad \langle z, \;\nabla f(x)\rangle \geq 0\},$$

$$Z_2(x) := \{z \in X : \langle z, \nabla g_\alpha(x) \rangle \geq 0 \quad \forall\, \alpha \in I_0(x),$$
$$\langle z, \nabla h_\beta(x) \rangle = 0 \quad \forall\, \beta \in J, \quad \langle z, \nabla f(x) \rangle < 0 \},$$

Proposition 2.1 *Let* $K = \nabla A_0(x^*) \cup \nabla B(x^*)$. *Assume*

$A1 :$ K *is compact;*

$A2 :$ $C(K)$ *is closed.*

If x^* *is a feasible point of problem (NLP), then the following are equivalent:*

(i) $Z_2(x^*) = \phi$.

(ii) There exist totally finite measures u^* *on* $(I, \mathbf{I})$ *and* λ^* *on* $(J, \mathbf{J})$ *such that*

   (a) $\ell'(x^*, u^*, \lambda^*) = 0$,

   (b) $g_\alpha(x^*) \geq 0 \quad \forall\, \alpha \in I$,

   (c) $h_\beta(x^*) = 0 \quad \forall\, \beta \in J$,

   (d) $u^*(I') = 0 \quad \forall\, I'$ *measurable* $\subset I \sim I_0(x^*)$,

   (e) $u^*(I') \geq 0 \quad \forall\, I'$ *measurable* $\subset I$.

We will refer to the conditions (a)–(e) in (ii) as the generalized first-order conditions.

Proof: Assume (ii) and suppose that $z \in Z_2(x^*)$. Then

$$
\begin{aligned}
0 &> \langle z, \nabla f(x^*) \rangle \\
&= \langle z, \int_I \nabla g_\alpha(x^*) u^*(d\alpha) - \int_J \nabla h_\beta(x^*) \lambda^*(d\beta) \rangle \\
&= \int_I \langle z, \nabla g_\alpha(x^*) \rangle u^*(d\alpha) - \int_J \langle z, \nabla h_\beta(x^*) \rangle \lambda^*(d\beta) \\
&\geq 0 ,
\end{aligned}
$$

which is a contradiction. This proves that (ii) implies (i).

Conversely, suppose that $Z_2(x^*) = \phi$. Then, if z satisfies

$$\langle z, \nabla g_\alpha(x^*) \rangle \geq 0 \qquad \forall\, \alpha \in I_0(x^*)$$
$$\langle z, \nabla h_\beta(x^*) \rangle \geq 0 \qquad \forall\, \beta \in J$$
$$\langle z, -\nabla h_\beta(x^*) \rangle \geq 0 \qquad \forall\, \beta \in J ,$$

z must also satisfy $\langle z, \nabla f(x^*)\rangle \geq 0$. But this implication is (i) in Lemma 2.1, so we may conclude that

$$
\begin{aligned}
\nabla f(x^*) &= \int_{\nabla A_0(x^*)} y \nabla F_0(dy) + \int_{\nabla B(x^*)} y \nabla F'(dy) - \int_{\nabla B(x^*)} y \nabla F''(dy) \\
&= \int_{I_0(x^*)} \nabla g_\alpha(x^*) t_0(d\alpha) + \int_J \nabla h_\beta(x^*) t'(d\beta) - \int_J \nabla h_\beta(x^*) t''(d\beta) \ .
\end{aligned}
$$

We now obtain conditions (a)–(e) by setting $u^* = t_0$ on $I_0(x^*), u^* = 0$ on $I \sim I_0(x^*)$, and $\lambda^* = -(t' - t'')$ on J. $\qquad\square$

Our statement of first-order condition (a) is somewhat nontraditional. Suppose that the g_α and h_β are elements of a real Hilbert space Γ. Assuming that the indicated expectations exist (which, of course, they may not), define the generalized Lagrangian function to be

$$
\ell(x, u, \lambda) := f(x) - \int_I g_\alpha(x) u(d\alpha) + \int_J h_\beta(x) \lambda(d\beta) \ .
$$

To conform to common practice, we would write condition (a) as

$$
\nabla_x \ell(x^*, u^*, \lambda^*) = \ell'(x^*, u^*, \lambda^*) = 0 \ .
$$

The following result establishes circumstances in which this representation is legitimate.

Proposition 2.2 *Fix $x \in X$. Let u and λ denote totally finite measures on $(I, \mathbf{I})$ and $(J, \mathbf{J})$. Assume that the expectations $\bar{g} := \int_I g_\alpha u(d\alpha)$ and $\bar{h} := \int_J h_\beta \lambda(d\beta)$ both exist. If the sets of functions $A := \{g_\alpha : \alpha \in I\}$ and B are each uniformly Lipschitz continuous, then $\nabla_x \ell(x, u, \lambda) = \ell'(x, u, \lambda)$.*

Proof: We must establish that

$$
\nabla \int_I g_\alpha(x) u(d\alpha) = \nabla \bar{g}(x) = \int_{\nabla A(x)} y \nabla F(dy) = \int_I \nabla g_\alpha(x) u(d\alpha) \ , \qquad (1)
$$

where ∇F is the measure on X induced by u. Clearly it suffices to prove that

$$
\langle \eta, \nabla \bar{g}(x) \rangle = \langle \eta, \int_{\nabla A(x)} y \nabla F(dy) \rangle \qquad \forall \ \eta \in X \ . \qquad (2)
$$

12

We note that

$$
\begin{aligned}
\langle \eta, \nabla \bar{g}(x) \rangle &= \bar{g}'(x)(\eta) \\
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \{ \bar{g}(x + \epsilon \eta) - \bar{g}(x) \} \\
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \{ \int_I [g_\alpha(x + \epsilon \eta) - g_\alpha(x)] u(d\alpha) \} \\
&= \lim_{\epsilon \to 0} \int_I \phi_\epsilon(x) u(d\alpha)
\end{aligned}
$$

and that

$$
\begin{aligned}
\langle \eta, \int_{\nabla A(x)} y \nabla F(dy) \rangle &= \int_{\nabla A(x)} \langle \eta, y \rangle \nabla F(dy) \\
&= \int_I \langle \eta, \nabla g_\alpha(x) \rangle u(d\alpha) \\
&= \int_I g_\alpha'(x)(\eta) u(d\alpha) \\
&= \int_I \lim_{\epsilon \to 0} \frac{1}{\epsilon} [g_\alpha(x + \epsilon \eta) - g_\alpha(x)] u(d\alpha) \\
&= \int_I \lim_{\epsilon \to 0} \phi_\epsilon(\alpha) u(d\alpha) \,.
\end{aligned}
$$

Since the g_α's are uniformly Lipschitz continuous, i.e. $\exists\ M < \infty$ such that $|g_\alpha(y) - g_\alpha(z)| \le M \|y - z\| \ \forall\ x, y \in X$ and $\forall\ \alpha \in I$, we have

$$
|\phi_\epsilon(\alpha)| \le \frac{M}{\epsilon} \|\epsilon \eta\| = M \|\eta\| < \infty \,.
$$

Then, since u is totally finite, we can apply the Dominated Convergence Theorem to interchange $\lim_{\epsilon \to 0}$ and $\int_I$. This establishes (2); hence, (1). The identical argument establishes that

$$
\nabla \int_J h_\beta(x) \lambda(d\beta) = \int_J \nabla h_\beta(x) \lambda(d\beta) \,,
$$

and the result follows. □

We now return to problem (NLP). As in the finite-dimensional case, in order to derive a necessity condition from Proposition 2.1, we must supplement the first-order conditions with a constraint qualification.

Definition 2.2 *Suppose that x^* is a feasible point of problem (NLP). We say that x^* satisfies the constraint qualification for problem (NLP) if:*

for each nonzero $z \in X$ satisfying $\langle z, \nabla g_\alpha(x^*) \rangle \geq 0 \; \forall \, \alpha \in I_0(x^*)$ and $\langle z, \nabla h_\beta(x^*) \rangle = 0 \; \forall \, \beta \in J$, there exists $\tau > 0$ and a continuous arc $C : [0, \tau) \to X$ satisfying

$$C(0) = x^* ,$$
$$C'(0) = z ,$$
$$g_\alpha(C(t)) \geq 0 \quad \forall \; t \in [0, \tau) \text{ and } \forall \, \alpha \in I ,$$
$$h_\beta(C(t)) = 0 \quad \forall \; t \in [0, \tau) \text{ and } \forall \, \beta \in J .$$

Our main result now follows precisely as in the finite-dimensional case.

Theorem 2.1 *Let $K = \nabla A_0(x^*) \cup \nabla B(x^*)$. Assume*
- *A1: K is compact;*
- *A2: $C(K)$ is closed.*

If x^* satisfies the constraint qualification for problem (NLP), then a necessary condition for x^* to be a local solution of problem (NLP) is that the first-order conditions hold.

Proof: We invoke Proposition 2.1. Suppose that x^* is a local solution and that $z \in Z_2(x^*)$. Clearly it must be that $z \neq 0$, so there exists a feasible continuous arc $C : [0, \tau) \to X$. Since x^* is a local solution, for $t > 0$ sufficiently small it must be that

$$f(C(t)) - f(C(0)) \geq 0$$

and therefore

$$\frac{1}{t}[f(C(t)) - f(C(0))] \geq 0 .$$

But this implies that

$$[f \circ C]'(0) = \langle \nabla f(C(0)), C'(0) \rangle = \langle \nabla f(x^*), z \rangle \geq 0 ,$$

which is a contradiction. $\qquad\qquad\square$

14

3 Discussion of Hypotheses

Let us now examine assumptions A1 and A2 of Theorem 2.1. To begin, recall that they automatically hold in the case of finite programming. Specifically, a finite set is compact, and it is well-known that a finitely generated cone is closed. Hence, assumptions A1 and A2 are exactly the price one must pay to extend the Farkas lemma approach to necessity conditions from finite programming to infinite programming. Of course this extension will be of no value if we cannot find reasonable conditions that imply assumptions A1, A2, and the constraint qualification. Such conditions are the subject of the present section; in the next section, we will present a meaningful example that satisfies our conditions.

We first consider assumption A1, the compactness assumption for the sets $\nabla A_0(x^*)$ and $\nabla B(x^*)$. Since problem (NLP) is stated without reference to these sets, it is obviously cumbersome to check hypotheses involving them. Fortunately, many problems will not require this.

Lemma 3.1 *Suppose that the g_α and h_β are elements of a real Hilbert space Γ. Fix $x \in X$ and let V_x denote evaluation at x. Suppose that the g_α and h_β are uniformly continuous and that V_x is a continuous functional on A and B. Assume that the index sets I and J have been topologized. If I and J are compact and the index maps $\alpha \mapsto g_\alpha$ and $\beta \mapsto h_\beta$ are continuous, then the sets $\nabla A_0(x)$ and $\nabla B(x)$ are compact.*

Proof: We argue in terms of the g_α. Given a sequence $\{\alpha_n\} \subset I_0(x)$, we claim that there exists $\alpha_0 \in I_0(x)$ and a subsequence $\{\alpha_{n'}\}$ such that $\nabla g_{\alpha_{n'}}(x) \to \nabla g_{\alpha_0}(x)$.

The indexing assumptions imply that A is compact. Since V_x is continuous on A, it follows that the level set $A_0(x) = \{g_\alpha : g_\alpha(x) = 0\} = \{g_\alpha : V_x(g_\alpha) = 0\}$ is closed, hence compact itself. Therefore there exists $\alpha_0 \in I_0(x)$ and a subsequence $\{\alpha_{n'}\}$ such that $g_{\alpha_{n'}} \to g_{\alpha_0}$.

The convergence indicated is in norm. However, since the g_α are uniformly continuous, the convergence must also be uniform. But this allows us to write

$$\lim_{n' \to \infty} \nabla g_{\alpha_{n'}}(x) = \nabla \lim_{n' \to \infty} g_{\alpha_{n'}}(x) = \nabla g_{\alpha_0}(x).$$

$\square$

We now derive conditions which imply that assumption A2 holds. We first derive a technical lemma about expectations that will be used to show that K compact and $0 \notin C_1(K)$ implies A2. This lemma derives from probability theory. An excellent reference for the requisite material is Billingsley [2].

Lemma 3.2 *Let H denote a real Hilbert space with inner product $(\cdot,\cdot)$. Let $\mathbf{M}_1(K)$ denote the family of probability measures that concentrate on the set $K \subset H$. If K is compact, then the set of expectations $C_1(K) := \{\int x\mu(dx) : \mu \in \mathbf{M}_1(K)\}$ is convex and compact.*

Remark: As mentioned before, the set $C_1(K)$ is essentially the convex hull of K.

Proof: Since K is compact, $\int (y,x)\mu(dx) \leq \|y\| \int \|x\| \mu(dx) \leq \|y\| \sup_{x \in K} \|x\| < \infty$; we are therefore assured that the expectations exist. The convexity of $C_1(K)$ follows immediately from the linearity of expectation.

To demonstrate compactness, consider the sequence $\{x_n = \int x\mu_n(dx) : \mu_n \in \mathbf{M}_1(K)\}$. Since K is compact, $\mathbf{M}_1(K)$ is tight. It follows from Prohorov's Theorem that there exists a weakly convergent subsequence of $\{\mu_n\}$, i.e. that there exists $\mu_0 \in \mathbf{M}_1(K)$ and a subsequence $\{\mu_{n'}\}$ such that

$$\int \phi(x)\mu_{n'}(dx) \to \int \phi(x)\mu_0(dx)$$

for all bounded continuous functions $\phi : H \to (-\infty, +\infty)$. Since K is compact, $(y, \cdot)$ is such a function; hence

$$\int (y,x)\mu_{n'}(dx) \to \int (y,x)\mu_0(dx) \quad \forall y \in H .$$

Then it must be that the Riesz representers

$$x_{n'} = \int x\mu_{n'}(dx) \to x_0 := \int x\mu_0(dx) ,$$

so the arbitrary sequence $\{x_n\}$ has a convergent subsequence. $\square$

We now remove the restriction that the positive measures used to form expectations have a total mass of unity.

Lemma 3.3 *Let H denote a real Hilbert space with inner product $(\cdot,\cdot)$ and origin 0. Let $\mathbf{M}(K)$ denote the family of totally finite, positive measures that concentrate on the set $K \subset H$. If K is compact and $0 \notin C_1(K)$, then $C(K) := \{\int x\mu(dx) : \mu \in \mathbf{M}(K)\}$ is convex and closed.*

Remark: As mentioned before, the set $C(K)$ is essentially the half-cone generated by the convex hull of K.

Remark: The conditions that K is compact and $0 \notin C_1(K)$ are sufficient but not necessary for the conclusion. To illustrate, let S be a closed subspace of H and let $K \subset S$ be any set such that 0 is an interior point of $C_1(K) \subset S$ relative to S, e.g. $\{x \in S : \|x\| < 1\}$. Then $C(K) = S$ is automatically convex and closed. However, the simple conditions stated in the lemma have a natural analog in the finite-dimensional theory and are entirely adequate for the example of Section 4.

Proof: Writing $C(K) = \{rx : x \in C_1(K), r = [0, +\infty)\}$, it follows from the convexity of $C_1(K)$ that $C(K)$ is a convex half-cone. We claim that $C(K)$ is also closed.

Toward that end, suppose that $\{y_n\} \subset C(K)$ with $\|y_n - \bar{y}\| \to 0$. Write $y_n = r_n x_n$ with $x_n \in C_1(K)$. By the compactness of $C_1(K)$, $\{x_n\}$ contains a subsequence $\{x_{n'}\}$ with $\|x_{n'} - \bar{x}\| \to 0$ for some $\bar{x} \in C_1(K)$. Moreover, since $0 \notin C_1(K)$, $\|\bar{x}\| > 0$.

Now let $\epsilon > 0$ be arbitrary. By construction, there exists $N(\epsilon)$ such that $n' \geq N(\epsilon)$ entails $\|x_{n'} - \bar{x}\| \leq \epsilon$ and $\|r_{n'} x_{n'} - \bar{y}\| \leq \epsilon$. It follows that, if $\epsilon < \|\bar{x}\|$ and $n' \geq N(\epsilon)$, then

$$\frac{\|\bar{y}\| - \epsilon}{\|\bar{x}\| + \epsilon} \leq r_{n'} \leq \frac{\|\bar{y}\| + \epsilon}{\|\bar{x}\| - \epsilon} \quad ,$$

so that $r_{n'} \to \bar{r} := \|\bar{y}\|/\|\bar{x}\|$. Hence,

$$\begin{aligned}
\|y_{n'} - \bar{r}\bar{x}\| &= \|r_{n'} x_{n'} - r_{n'}\bar{x} + r_{n'}\bar{x} - \bar{r}\bar{x}\| \\
&\leq r_{n'}\|x_{n'} - \bar{x}\| + |r_{n'} - \bar{r}| \, \|\bar{x}\| \\
&\to 0 .
\end{aligned}$$

By the uniqueness of limits, $\bar{y} = \bar{r}\bar{x} \in C(K)$. $\square$

Notice that the hypothesis that $0 \notin C_1(K)$ is closely related to the oft-imposed (in finite programming) condition of regularity. A feasible point x^* is said to be *regular* if the set K is linearly independent, i.e. if no finite nonzero linear combination of the constraint gradients at x^* can vanish. Our condition is somewhat stronger in one respect, but much weaker in another. On the one hand, we consider arbitrary measures (weights) on K, not just

17

finitely supported ones. This is analogous to infinite linear combinations, hence stronger; on the other hand, we only consider probability measures (nonnegative weights totalling unity); this is analogous to convex combinations instead of linear combinations, hence weaker.

In finite programming, if x^* is a regular point, then x^* must satisfy the constraint qualification. This pleasant property does not hold in infinite programming; in fact, since the number of linearly independent gradients cannot exceed the dimension of the space X, the notion of regularity is wholly inappropriate for the case of semi-infinite programming and somewhat inappropriate in the case of infinite programming. Accordingly, we will search for other conditions that will imply the constraint qualification.

The simplest situation is the one in which all of the constraints are linear. If x^* and z are as in Definition 2.2, then the arc $C(t) = x^* + tz$ satisfies

$$
\begin{aligned}
C(0) &= x^* ; \\
C'(0) &= z : \\
g_\alpha(C(t)) &= 0 \quad \forall\, t \geq 0, \quad \forall\, \alpha \in I_0(x^*) ; \\
h_\beta(C(t)) &= 0 \quad \forall\, t \geq 0, \quad \forall\, \beta \in J .
\end{aligned}
$$

Moreover, for each $\alpha \in I \sim I_0(x^*)$ (the nonbinding constraints), there exists $\tau(\alpha) > 0$ such that

$$
g_\alpha(C(t)) \geq 0 \quad \forall\, t \in [0, \tau(\alpha)).
$$

If the number of nonbinding constraints is finite, then we can take $\tau = \inf_\alpha \{\tau(\alpha)\} > 0$ and the constraint qualification is automatically satisfied. Otherwise, it may be that $\inf_\alpha \{\tau(\alpha)\} = 0$ and the constraint qualification may not hold. We are therefore content to establish that the constraint qualification holds for one important family of examples.

Both control theory and statistics abound with constraints of the sort that a function be bounded by certain values. The following result addresses the prototypical case; we hope that the method of proof will suffice for a variety of applications.

Lemma 3.4 *Let X denote a real Hilbert space of functions $x : I \to (-\infty, +\infty)$. Let g_α denote evaluation at $\alpha \in I$. If X is a proper functional Hilbert space, i.e. if the g_α are continuous, then the collection of inequality constraints*

$$
g_\alpha(x) = x(\alpha) \geq 0 \quad \forall\, \alpha \in I
$$

18

satisfies the constraint qualification.

Proof: Since the g_α are continuous and linear, $\nabla g_\alpha(x)$ exists $\forall\, x \in X$. Suppose that $x^* \in X$ is a feasible point and that a nonzero $z \in X$ satisfies

$$\langle z, \nabla g_\alpha(x^*)\rangle = g'_\alpha(x^*)(z) = z(\alpha) \geq 0 \quad \forall\, \alpha \in I_0(x^*),$$
$$\text{i.e. } \forall\, \alpha \text{ such that } x^*(\alpha) = 0 .$$

For $t > 0$, let

$$K(t) := \{y \in X : x^*(\alpha) + ty(\alpha) \geq 0 \quad \forall\, \alpha \in I\} .$$

We note that the sets $K(t)$ are nested, for suppose that $t_0 < t_1$. If $0 \leq x^*(\alpha) + t_1 y(\alpha) \quad \forall\, \alpha \in I$, i.e. $y \in K(t_1)$, then

$$0 \leq \frac{t_0}{t_1} x^*(\alpha) + t_0 y(\alpha) \leq x^*(\alpha) + t_0 y(\alpha) \quad \forall\, \alpha \in I ,$$

i.e. $y \in K(t_0)$. Thus, $K(t_1) \subset K(t_0)$.

Suppose that $t_n \downarrow t_0$. Then $\cup_n K(t_n) \subset K(t_0)$; hence, the closure of $\cup_n K(t_n)$ is contained in the closure of $K(t_0)$, which is just $K(t_0)$. To establish the converse, let $y_0 \subset K(t_0)$ and set $y_n = (t_0/t_n)y_0$. Then $y_n \subset K(t_n)$ and $\|y_n - y_0\| \to 0$; hence, $K(t_0)$ is contained in the closure of $\cup_n K(t_n)$.

Suppose that $t_n \uparrow t_0$. Then $\cap_n K(t_n)$, i.e.

$$\{x^*(\alpha) + t_n y(\alpha) \geq 0 \quad \forall \alpha \in I\} \quad \forall t_n .$$

Then

$$\{x^*(\alpha) + t_0 y(\alpha) \geq 0 \quad \forall \alpha \in I\} ,$$

i.e. $y \in K(t_0)$.

It now follows from Proposition 3.22 in Attouch [1] that, if $t_n \to t_0$, then $K(t_n) \to K(t_0)$ in the sense of Mosco-convergence of closed convex sets. Furthermore, Proposition 3.34 in Attouch [1] states that the Mosco-convergence of closed convex sets is equivalent to the convergence of the projections of an arbitrary point into these sets. Therefore, let $z^*(t)$ denote the projection of Z into $K(t)$. Then, $t_n \to t_0$ entails $\|z^*(t_n) - z^*(t_0)\| \to 0$, and we conclude that $z^*(t)$ is a continuous arc for $t > 0$. Moreover, since $x^*(\alpha) = 0$ entails $z(\alpha) \geq 0$, z is contained in the closure of $\cup_{t>0} K(t)$. We can therefore close the arc by setting $z^*(0) = z$.

Now let $C(t) := x^* + tz^*(t)$. By construction, C is a feasible continuous arc with $C(0) = x^*$. Moreover,

$$\lim_{t \to 0} \frac{1}{t} \|C(t) - C(0) - tz\| = \lim_{t \to 0} \frac{1}{t} \|tz^*(t) - tz\| = \lim_{t \to 0} \|z^*(t) - z\| = 0 ,$$

so $C'(0) = z$. This verifies the conditions specified by Definition 2.2. □
Remark: It is also possible to give an elementary proof that the arc $z^*(t)$ is continuous. The use of Mosco-convergence was suggested to us anonymously by the referee. The equivalence of Mosco-convergence and the convergence of the projections of an arbitrary point is due to Sonntag [18].

4   An Example

We now apply our results to obtain necessity conditions for a well-known problem from the statistical literature on probability density estimation. Watson and Leadbetter [24] sought to minimize the mean integrated squared error of a kernel probability density estimator. Specifically, given independent and identically distributed random variables $X_1, \ldots, X_n$ with probability density function δ, they analyzed the optimization problem

$$\operatorname*{minimize}_{K_n \in L^2(-\infty, +\infty)} E \int_{-\infty}^{\infty} \left[\frac{1}{n} \sum_{i=1}^{n} K_n(x - X_i) - \delta(x) \right]^2 dx .$$

It turns out that solutions are typically not everywhere nonnegative, which results in estimates that are not themselves probability densities. This is a matter of taste, but if we prefer to estimate densities with densities, then we must confront the constrained optimization problem

$$
\begin{aligned}
&\operatorname*{minimize}_{K_n \in L^2(-\infty, +\infty)} && E \int_{-\infty}^{\infty} \left[\frac{1}{n} \sum_{i=1}^{n} K_n(x - X_i) - \delta(x) \right]^2 dx \\
&\text{subject to} && K_n(x) \geq 0 \quad x \in (-\infty, +\infty) \\
& && \int_{-\infty}^{\infty} K_n(x) dx = 1 .
\end{aligned}
$$

This problem does not yield to variational methods, making it a natural candidate for the application of our multiplier theory. We proceed to formulate it in that context.

Consider the Sobolev space $H^1[\alpha_1, \alpha_2]$, which is defined by endowing the vector space

$$\{x : x^{(j)} \in L^2[\alpha_1, \alpha_2] \text{ for } j = 0, 1\}$$

with the inner product

$$\langle x, y \rangle = \sum_{j=0}^{1} \langle x^{(j)}, y^{(j)} \rangle_{L^2[\alpha_1, \alpha_2]}.$$

It should be noted that the derivatives in the definition of $H^1[\alpha_1, \alpha_2]$ are taken in the sense of distributions. It is well known that $H^1[\alpha_1, \alpha_2]$ is a proper functional Hilbert space and that each element of $H^1[\alpha_1, \alpha_2]$ is absolutely continuous. See Appendix I of Tapia and Thompson [20] for a discussion of the analogous Sobolev space, $H^1(-\infty, +\infty)$. Notice that, if $\delta \in H^1(-\infty, +\infty)$, then the restriction of δ to $[\alpha_1, \alpha_2]$ is an element of $H^1[\alpha_1, \alpha_2]$.

We now return to the problem of Watson and Leadbetter, which we reformulate as problem (WL):

$$
\begin{array}{ll}
\underset{x_n \in X}{\text{minimize}} & f(x_n) = E \int_{-\infty}^{\infty} \left[\frac{1}{n} \sum_{i=1}^{n} \bar{x}_n(\alpha - \chi_i) - \delta(\alpha) \right]^2 d\alpha \\[2ex]
\text{(WL)} \quad \text{subject to} & g_\alpha(x_n) = x_n(\alpha) \geq 0 \ \forall \, \alpha \in I \\[2ex]
& h(x_n) = \int_I x_n(\alpha) d\alpha - 1 = 0 \; ,
\end{array}
$$

where $I = [\alpha_1, \alpha_2]$, $X = H^1[\alpha_1, \alpha_2]$, and $\bar{x}_n$ denotes the extension of x_n to $(-\infty, +\infty)$ defined by $\bar{x}_n(\alpha) = 0$ if $\alpha \notin I$; and where the expectation is taken with respect to the independent and identically distributed random variables $\chi_i, i = 1, \ldots, n$, having probability density function $\delta \in H^1(-\infty, +\infty)$. We have modified the original problem in two ways. First, we have demanded some additional smoothness. Second, we have restricted attention to kernels supported on $[\alpha_1, \alpha_2]$. We proceed to verify that Theorem 2.1 can be applied to problem (WL).

The point evaluation functionals $g_\alpha \in \Gamma = X^*$ are both linear and continuous, hence continuously differentiable and also uniformly continuous. It is also easily checked that $f, h \in C^1(X)$. Furthermore, the set $\nabla B(x) = \{\nabla h(x)\}$ is obviously compact. We also have

Lemma 4.1 *For problem (WL), the set $\nabla A(x) = \{\nabla g_\alpha(x) : \alpha \in I\}$ is compact.*

Proof: We apply Lemma 3.1. The point evaluation functionals $V_x \in \Gamma^*$ are continuous, since $V_x(g_\alpha) = g_\alpha(x) = x(\alpha)$. Since I is compact, it remains only to demonstrate that the index map $\alpha \mapsto g_\alpha$ is continuous.

Consider the optimization problem

$$\underset{x}{\text{minimize}} \qquad \int_{a_1}^{a_2} [x'(\alpha)]^2 d\alpha$$

$$\text{subject to} \qquad x(a_1) = b_1, x(a_2) = b_2 .$$

It is a trivial exercise in the calculus of variations to establish that the minimizer is a straight line with slope

$$x'(\alpha) = (b_2 - b_1)/(a_2 - a_1).$$

This yields a minimum objective function value of $|b_2 - b_1|^2/|a_2 - a_1|$. It follows that any $x \in X$ with $x(a_1) = b_1$ and $x(a_2) = b_2$ must satisfy

$$\|x\|^2 \geq |b_2 - b_1|^2/|a_2 - a_1|. \tag{3}$$

Now suppose that $\alpha_n \to \alpha_0$ as $n \to \infty$. Then (3) allows us to write

$$\|g_{\alpha_n} - g_{\alpha_0}\| = \sup_{\|x\| \leq 1} |g_{\alpha_n}(x) - g_{\alpha_0}(x)| = \sup_{\|x\| \leq 1} |x(\alpha_n) - x(\alpha_0)|$$

$$\leq \sup_{\|x\| \leq 1} |\alpha_n - \alpha_0|^{\frac{1}{2}} \|x\| = |\alpha_n - \alpha_0|^{\frac{1}{2}} \to 0 \text{ as } n \to \infty . \quad \Box$$

Next, we show that our conditions on K hold.

Lemma 4.2 *For problem (WL), let $K = \nabla A(x) \cup \nabla h(x)$. Then $C_1(K)$ does not contain the origin of $X = H^1[\alpha_1, \alpha_2]$.*

Proof: We exploit the fact that the gradient is the Riesz representer of the directional derivative. Let $\eta \in X$; then

$$g_\alpha'(x)(\eta) = \lim_{t \to 0} \frac{1}{t} [g_\alpha(x + t\eta) - g_\alpha(x)]$$

$$= \lim_{t \to 0} \frac{1}{t} [x(\alpha) + t\eta(\alpha) - x(\alpha)]$$

$$= \eta(\alpha)$$

22

and

$$h'(x)(\eta) = \lim_{t \to 0} \frac{1}{t}[h(x + t\eta) - h(x)]$$

$$= \lim_{t \to 0} \frac{1}{t} \left[\int_I [x + t\eta](\alpha)d\alpha - 1 - \int_I x(\alpha)d\alpha + 1 \right]$$

$$= \int_I \eta(\alpha)d\alpha .$$

Hence, $\nabla g_\alpha(x)$ must satisfy

$$\langle \nabla g_\alpha(x), \eta \rangle = \eta(\alpha) \quad \forall \, \eta \in X$$

and $\nabla h(x)$ must satisfy

$$\langle \nabla h(x), \eta \rangle = \int_I \eta(\alpha)d\alpha \quad \forall \, \eta \in X .$$

Now suppose that there exists $\mu \in \mathbf{M}_1(K)$ such that $\int_K y\mu(dy) = 0$. Let $\lambda = \mu(\nabla h(x))$ and let $(1 - \lambda)u$ denote the measure on $(I, \mathbf{I})$ induced by μ. Then it must be that, $\forall \, \eta \in X$,

$$0 = \langle 0, \eta \rangle$$

$$= \langle \int_K y\mu(dy), \eta \rangle$$

$$= \int_K \langle y, \eta \rangle \mu(dy)$$

$$= \int_{\nabla A(x)} \langle y, \eta \rangle \mu(dy) + \lambda \langle \nabla h(x), \eta \rangle$$

$$= \int_{\nabla A(x)} \eta(\alpha)\mu(dy) + \lambda \int_I \eta(\alpha)d\alpha$$

$$= (1 - \lambda) \int_I \eta(\alpha)u(d\alpha) + \lambda \int_I \eta(\alpha)d\alpha . \tag{4}$$

But the last expression in (4) is strictly positive if $\eta \in X$ is strictly positive on I; hence, $C_1(K)$ cannot contain the origin of X. $\qquad \square$

Remark: If u is a finitely supported signed measure, say $u = \sum_{i=1}^m u_i \mathbf{1}(\alpha_i)$, where $\mathbf{1}$ denotes point-mass, then (4) reduces to

$$0 = (1 - \lambda) \sum_{i=1}^m u_i \eta(\alpha_i) + \lambda \int_I \eta(\alpha)d\alpha .$$

If $\lambda = 1$, this equality fails for (say) $\eta(\alpha) \equiv 1$; if $\lambda \neq 1$, this equality fails for any η satisfying $\eta(\alpha_i) = -u_i$ and $\int_I \eta(\alpha)d\alpha = 0$. Thus, the condition of regularity also holds for problem (WL). Notice, however, that the restriction to finite linear combinations in the definition of linear independence is crucial to this conclusion. If arbitrary signed measures are allowed, then take u to be the negative uniform measure on I and put $\lambda = 1/(\alpha_2 - \alpha_1 + 1)$. Then the last expression in (4) is

$$(1 - \lambda) \int_I \eta(\alpha)u(d\alpha) + \lambda \int_I \eta(\alpha)d\alpha$$

$$= \left[1 - \frac{1}{\alpha_2 - \alpha_1 + 1}\right] \frac{1}{\alpha_2 - \alpha_1} \int_I \eta(\alpha)d\alpha + \frac{1}{\alpha_2 - \alpha_1 + 1} \int_I \eta(\alpha)d\alpha ,$$

which does indeed vanish $\forall \eta \in X$. This distinction should not be surprising. Roughly stated, finitely many values do not determine a function's Lebesgue integral, but all values together do.

Finally, the equality constraint in problem (WL) is easily incorporated into the proof of Lemma 3.4. This provides a means of verifying that any feasible point for problem (WL) satisfies the constraint qualification. Theorem 2.1 therefore applies: a necessary condition for x_n^* to be a local solution of problem (WL) is that the first-order conditions hold.

Let us make some further observations concerning problem (WL). The objective function is strictly convex and the constraint set is convex. It follows that any local solution will be the unique global solution. It is well known that the variational inequality which serves as a necessity condition when the constraint set is convex serves as a sufficiency condition when the objective function is also convex. A rather straightforward argument can be used to show that, in the case of a convex constraint set, condition (i) of Proposition 2.1, namely $Z_2(x^*) = \phi$, implies the variational inequality necessity condition. These comments say that, in the case of a convex program where the constraint qualification holds (as is the case for problem (WL)), the existence of Lagrange multipliers (Proposition 2.1) is both necessary and sufficient for x^* to be a global minimizer.

Our theory, the above comments and some straightforward computations lead us to the following result concerning problem (WL): x_n^* is the unique global minimizer if and only if there exists a totally finite measure concentrating on $[\alpha_1, \alpha_2]$, with density function u_n^*, and a real number λ_n^*, such that

(a) $u_n^*(\alpha) = 2\frac{n-1}{n}[x_n^* * \delta * \bar\delta](\alpha) - 2[\delta * \bar\delta](\alpha) + \frac{2}{n}x_n^*(\alpha) + \lambda_n^*$ $\forall\, \alpha \in [\alpha_1, \alpha_2]$,

(b) $x_n^*(\alpha) \geq 0$ $\forall\, \alpha \in [\alpha_1, \alpha_2]$,

(c) $\int_{\alpha_1}^{\alpha_2} x_n^*(\alpha)d\alpha = 1$,

(d) $x_n^*(\alpha)u_n^*(\alpha) = 0$ $\forall\, \alpha \in [\alpha_1, \alpha_2]$,

(e) $u_n^*(\alpha) \geq 0$ $\forall\, \alpha \in [\alpha_1, \alpha_2]$.

In condition (a), $\bar\delta(\alpha) := \delta(-\alpha)$, and $*$ denotes convolution.

Since problem (WL) is highly nontrivial, it is not surprising that the corresponding necessity conditions are somewhat complicated. A more detailed analysis of these conditions was undertaken by Trosset [21]. Nevertheless, it is evident from the material presented here that the theory developed in Sections 2 and 3 can be productively applied to a body of problems admitting an infinite programming formulation.

Acknowledgements

References

[1] H. Attouch. *Variational Convergence for Functions and Operators*. Pitman, Boston, 1984.

[2] P. Billingsley. *Convergence of Probability Measures*. John Wiley and Sons, New York, 1968.

[3] G. A. Bliss. Normality and abnormality in the calculus of variations. *Transactions of the American Mathematical Society*, 43:365–376, 1938.

[4] J. Dieudonné. Sur le théorème de Hahn-Banach. *Rev. Sci.*, 79:642–643, 1941.

[5] J. Farkas. Über die Theorie der einfachen Ungleichungen. *Journal für die reine und angewandte Mathematik*, 124:1–27, 1901.

[6] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, New York, 1968.

[7] I. V. Girsanov. *Lectures on Mathematical Theory of Extremum Problems*. Springer-Verlag, New York, 1972.

[8] A. Goberna, M. A. López, and J. Pastor. Farkas-Minkowski systems in semi-infinite programming. *Applied Mathematics and Optimization*, 7:295–308, 1981.

[9] M. Guignard. Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space. *SIAM Journal on Control*, 2:232–241, 1969.

[10] M. R. Hestenes. *Optimization Theory: The Finite Dimensional Case*. John Wiley and Sons, New York, 1975.

[11] F. John. Extremum problems with inequalities as subsidiary conditions. In K. O. Friedrichs, O. E. Neugebauer, and J. J. Stoker, editors, *Studies and Essays Presented to R. Courant on His 60th Birthday*, pages 187–204, Interscience Publishers, New York, 1948.

[12] W. Karush. *Minima of Functions of Several Variables with Inequalities as Side Conditions*. Master's thesis, University of Chicago, 1939.

[13] H. W. Kuhn. Nonlinear programming: a historical view. In R. W. Cottle and C. E. Lemke, editors, *Nonlinear Programming*, pages 1–26, SIAM-AMS, American Mathematical Society, Providence, 1976. An appendix summarizes Karush's master's thesis.

[14] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, University of California Press, Berkeley, 1951.

[15] A. Mayer. Begrüdung der Lagrange'schen Multiplicatorenmethode in der Variationsrechnung. *Mathematische Annalen*, 26:74–82, 1886.

[16] K. R. Parthasarathy. *Probability Measures on Metric Spaces*. Academic Press, New York, 1967.

[17] A. Prékopa. On the development of optimization theory. *American Mathematical Monthly*, 87:527–542, 1980.

[18] Y. Sonntag. Interprétation géométrique de la convergence d'une suite de convexes. *C. R. Acad. Sc. Paris, Série A*, 282:1099–1100, 1976.

[19] A. Takayama. *Mathematical Economics*. Dryden Press, Hinsdale, IL, 1974.

[20] R. A. Tapia and J. R. Thompson. *Nonparametric Probability Density Estimation*. Johns Hopkins University Press, Baltimore, 1978.

[21] M. W. Trosset. Optimal shapes for kernel density estimation. *Communications in Statistics — Theory and Methods*, 22:375–391, 1993.

[22] F. A. Valentine. *Convex Sets*. Krieger, New York, 1964. Reprinted with corrections in 1976.

[23] F. A. Valentine. The problem of Lagrange with differential inequalities as added side conditions. In *Contributions to the Calculus of Variations, 1933–1937*, pages 403–447, University of Chicago Press, Chicago, 1937. Ph.D. thesis, University of Chicago.

[24] G. S. Watson and M. R. Leadbetter. On the estimation of the probability density, I. *Annals of Mathematical Statistics*, 34:480–491, 1963.